Detecting Stock Calendar Effects in U.S. Census Bureau Inventory Series

Natalya Titova* Brian Monsell†

Abstract
U.S. Census Bureau retail, wholesale, and manufacturing inventory series are evaluated for the presence of stock trading day and stock Easter effects. We are especially interested in the detection of the one-coefficient stock trading day effect described in Findley and Monsell (2009) and a stock Easter effect described in Findley (2009). Using the diagnostic capabilities of X-13A-S, we utilize likelihood statistics, spectral analysis, and forecasting diagnostics to decide whether stock regressors should be included in the models of inventory series, as well as what type of stock regressor (full implementation or one-coefficient trading day, end-of-month stock versus choosing a sample day, etc.) Results of this study are presented and discussed.

Key Words: RegARIMA model, trading day adjustment, moving holiday, stock Easter effect, spectral diagnostics

1. Introduction

In this study, several U.S. Census Bureau end-of-month inventory series are analyzed for the presence of stock trading day and stock Easter effects. End-of-month inventory series are a type of stock economic time series which are a result of a sum of monthly inflows and outflows. These end-of-month stock series can be considered as monthly accumulations of flow series.

In addition, the minimum AICC criterion and other modeling diagnostics are utilized to select what type of stock trading day regressors should be included in the final models of the inventory series. Stock Easter moving holiday regressors developed in Findley (2009) are also considered.

2. Regressors (calendar effects) of interest

Since an end-of-month stock series can be viewed as an accumulation of consecutive monthly flow series values, a regressor for the stock series can be derived by accumulating values of an appropriate flow series regressor (see Bell (1984, 1995) for more details). The stock day (denoted by d) is the day in a month on which inventory is assumed to be taken. Stock trading day and Easter regressors for flow series from day $d+1$ of one month to day d of the next month would differ from the usual calendar month flow regressors. The flow Easter regressors within X-12-ARIMA assume that the level of activity changes on the w-th

*Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233, natalya.titova@census.gov
†Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233, brian.c.monsell@census.gov

day before the holiday for a specified w, and remains at the new level until the day before the holiday.

The following calendar effects are analyzed in the study (the X-13A-S arguments needed to produce these regressors are given in brackets, where appropriate):

1. Full end-of-month stock trading day (`tdstock[31]`);
2. One-coefficient end-of-month stock trading day (`tdstock1coef[31]`);
3. Full stock trading day with a stock day of 28 (`tdstock[28]`);
4. One-coefficient stock trading day with a stock day of 28 (`tdstock1coef[28]`);
5. End-of-month stock Easter$[w]$ with an effect window of $w = 1, 8, 15$ (`easterstock[w]`);
6. Stock Easter$[w]$ with a stock day of 28 and an effect window of $w = 1, 8, 15$.

The last regressor, the stock Easter$[w]$ with $d = 28$, was generated using the *genhol* utility (available from the Census Bureau's X-12-ARIMA website) that generates user-defined moving holiday regressors.

Stock trading day regressors are discussed in detail by Findley and Monsell (2009), and stock Easter regressors are described by Findley (2009). All regressors have been implemented in a developmental version of the X-13A-S program (it should be noted that X-13A-S uses functions of X-12-ARIMA, the earlier program.)

In any model an end-of-month stock trading day regressor should only be combined with an end-of-month stock Easter, not with a stock Easter with $d = 28$; similarly, a stock trading day with $d = 28$ regressor should only be combined with a stock Easter with $d = 28$.

Setting the stock day to 28 could be useful for some series because there can be respondents who take inventory on a different schedule (such as using a pattern of 4-4-5 week months throughout the year) and thus do not give true end-of-calendar month values. For series in which these respondents make a large contribution, regressors with $d = 28$ might be better than end-of-month regressors.

3. Series in the empirical study

For this empirical study, two sets of Census Bureau inventory series were analyzed.

A set of 27 series from the Monthly Wholesale and Retail Sales Report were used, out of which there were 7 U.S. retail inventory series and 20 U.S. wholesale inventory series. The data used started in January 1995 and ended in July 2008.

In addition, 96 inventory series of the U.S. Census Bureau's monthly U.S. Manufacturers' Shipments, Inventories and Orders Survey (the M3 Survey) were used in the study. These series will be referred to as the M3 series or manufacturing inventory series. The span of data used in this study ends in July 2008, but the starting date varied from January 1992 to January 1997, according to the choice made for regARIMA modeling for each series.

The final two code letters in names of M3 series are TI – Total Inventories; MI – Materials and Supplies Inventories; WI – Work in Process Inventories; FI – Finished Goods Inventories. The latter three are components of the TI series of the same category.

For information on data collection methods and reliability of the estimates, access the Economic Indicators page on the Internet. Program overviews and current data are available from links on that page (`http://www.census.gov/cgi-bin/briefroom/BriefRm`).

The regARIMA models used by the Census Bureau to produce the monthly seasonal adjustments for these series are the models used for this study. The only part of the models that is changed throughout this analysis is whether calendar regressors are used (or not used). For each series, the outlier set currently used for production was maintained, so all regARIMA models used for a given series had the same outlier regressors.

4. Steps of analysis

This section provides a general description of the procedures used to select models and to evaluate the modeling diagnostics used in the analysis of selected models.

4.1 The minimum AICC criterion in selecting models

Findley and Monsell (2009) used likelihood ratio (LR) tests to test whether or not a stock day-of-week regressor should be included in the regARIMA model for a series. Let ΔL denote the difference between the maximum Gaussian log-likelihood values for an unconstrained stationary time series model and for a model nested within it having ν fewer independent parameters. Under the null hypothesis that the nested model is correct, Taniguchi and Kakizawa (2000, p. 61) show that the asymptotic distribution of $2\Delta L$ is chi-square with ν degrees of freedom:

$$2\Delta L \sim \chi^2_\nu.$$

We thus pick the unconstrained model when $2\Delta L$ exceeds the 5% critical value for χ^2_ν.

Some model comparisons involve non-nested models: neither is a special case of the other. For these comparisons, models were chosen using differences in AICC (Akaike's Information Criterion, bias Corrected) values. AICC, also known as the F-corrected AIC, is a version of Akaike's Information Criterion (AIC) which contains a bias correction for series with a small sample size, see Hurvich and Tsai (1989). In general, a model with the smallest AICC value is preferred.

To define AICC, let L_N denote the log-likelihood function of a model, and n_p denote the number of estimated parameters in the model, including the white noise variance. Parameters of this model are estimated from time series data of length N obtained after applying the differencing polynomial of this model. Then $AICC_N$ is defined by:

$$AICC_N = -2L_N + 2n_p \left(1 - \frac{n_p + 1}{N}\right)^{-1}. \quad (1)$$

Taking the fact that $2\Delta L \sim \chi^2_\nu$, it follows that differences of AICC values can be used to form the LR tests for nested comparisons. Let $AICC_N^A$ and $AICC_N^B$ refer to AICC values of regARIMA models A and B, respectively. We assume that the model B is an extension of the model A and has ν additional estimated parameters, compared with the model A: $n_p^B = n_p^A + \nu$. Let $dAICC_N = AICC_N^A - AICC_N^B$. For large N we see that $dAICC_N \approx (-2L_N^A + 2n_p^A) - (-2L_N^B + 2n_p^B) = 2\Delta L - 2\nu$. At the 5% level of

Table 1: 5% *level of significance critical values* $\Delta_{0.05}$ *for models* A *and* B. The model B is preferred over the model A if $dAICC_N = AICC_N^A - AICC_N^B > \Delta_{0.05}$. In order to compare a model B with a model A that is a model with calendar regressor(s), both models B and A should have AICC values that are significantly lower than the AICC value of the model without calendar regressors.

Model A	**Model** B	ν	$\Delta_{0.05}$
No calendar regressor	Full end-of-month stock trading day	6	**0.592**
No calendar regressor	One-coefficient end-of-month stock trading day	1	**1.841**
No calendar regressor	Full stock trading day with $d = 28$	6	**0.592**
No calendar regressor	One-coefficient stock trading day with $d = 28$	1	**1.841**
One-coefficient end-of-month stock trading day	Full end-of-month stock trading day	5	1.120
One-coefficient stock trading day with $d = 28$	Full stock trading day with $d = 28$	5	1.120
Any end-of-month stock trading day	*Any* stock trading day with $d = 28$	–	2.0‡
Any stock trading day	*Same* stock trading day and *any* stock Easter	1	1.841
No calendar regressor	*Any* stock Easter	1	1.841

‡For this non-nested comparison, the value 2.0 given for $\Delta_{0.05}$ does not ensure statistical significance at the $\alpha = 0.05$ level but is a somewhat heuristic value suggesting a preponderance of support for model B, see Burnham and Anderson (2004, p. 271).

significance, we can determine the critical value $\Delta_{0.05}$ for $dAICC_N$ such that:

$$\lim_{N\to\infty} Pr\left(dAICC_N > \Delta_{0.05}\right) = Pr\left(\chi^2_\nu > 2\nu + \Delta_{0.05}\right) = 0.05,$$

using the asymptotic χ^2_ν distribution of $2\Delta L = 2\left(L_N^B - L_N^A\right) = AIC_N^A - AIC_N^B + 2\nu$.

Table 1 shows the values of $\Delta_{0.05}$ for the eight relevant nested comparisons. For example, to test the null hypothesis that the constrained version of the six-coefficient trading day model that defines the one-coefficient model is correct against the alternative that the full, unconstrained model is required, we use $\nu = 5$. Then $\Delta_{0.05} = 1.120$ is the minimum AICC difference required to reject the one-coefficient model in favor of the full model.

As noted above, some of the required regressor comparisons are non-nested. One case is the comparison of stock Easter regressors for pre-holiday intervals of different lengths w (e.g., $w = 8$ versus $w = 15$). For this case, the comparison of AICCs reduces to the log-likelihood comparison, as the "penalty terms" (the second term on the right-hand side of (1)) for all models are the same.

The second case is the comparison of stock regressors for different stock days (e.g., end-of-month stocks versus 28th-day stocks). One criterion was imposed to prefer a regressor for non-end-of-month stocks (e.g., for $d = 28$) over an end-of-month stock regressor for a series nominally consisting of end-of-month stocks: the AICC of the model with the non-end-of-month stock regressor must be smaller than the AICC of the model with the end-of-month stock regressor by at least 2.0.

In our study, trading day regressor decisions were made before Easter regressor decisions. The one-coefficient and the full trading day regressors were considered for end-of-month and 28th-day stocks, defining four regression models to be compared first to the model without regressors. Any stock trading day models that were preferred were then compared among themselves by using the criteria described above for nested and non-nested comparisons. In this way, two groups of series were formed: those without trading day regressor and those with a preferred trading day regressor.

The resulting models with an end-of-month stock trading day regressor were augmented with the end-of-month stock Easter variable `easterstock[w]` for interval lengths $w =$

$1, 8, 15$ taken one at a time. The same was done with 28th-day stocks ($d = 28$). Since $\Delta_{0.05} = 1.841$ when $\nu = 1$, an augmented model received further consideration only if its AICC was smaller than the AICC of the model without the Easter regressor by 1.841. When the model without Easter regressor was rejected in favor of more than one model with a stock Easter regressor, a model with a stock Easter regressor that has the smallest AICC value was usually preferred, but we also compared diagnostic results of these models in order to select the model which had relatively better diagnostics.

4.2 History diagnostics

All the diagnostics described in this section can be generated using the *history* spec of X-13A-S. When the optional *history* spec is specified by the user, X-13A-S will perform a sequence of runs for truncated versions of the time series to simulate the passage of time from month to month (or quarter to quarter), starting from a given time point. The program creates a historical record of certain statistics of interest for the span of data from the given starting point to the end of the series. The user can specify which records are to be collected (the choices include seasonal adjustment or trend revisions, out-of-sample forecast errors, and likelihood statistics) and the starting date of the historical record[1]. Various summary diagnostics of the statistics collected are produced, depending on the statistics chosen by the user.

4.2.1 Plot of differences of the AIC values

It can be useful to create a plot of the history of the differences of the AIC values of a model without calendar regressors and of a model with calendar regressor(s) over the last eight years. These series of AIC values are produced by the *history* spec of X-13A-S, where each observation is the AICC value for a given model for the span of data that ends at that observation. Henceforth, these plots will be called AIC history plots. The AIC history graph is discussed in Findley, Monsell, Bell, Otto and Chen (1998). We can look at the range of differences of the AICs — a range that is mostly in the positive region usually supports the model with calendar regressor(s) over the model without calendar regressor(s). Also, if the curve of AIC differences seems to be sloping upward over time, then there is a probability that the preference for the model with calendar regressor(s) grows over time. However, the absence of calendar effects is not indicated by a consistent movement of the curve.

4.2.2 Plot of differences of the sum of squared forecast errors

As it was already mentioned, the *history* spec of X-13A-S also generates differences of the accumulating sums of squared forecast errors between two models for forecast leads of interest (1 and 12 in our study). The user can specify what forecasts are to be generated (the default is one-step and one-year ahead forecasts.) These forecast errors can be saved and

[1] In our study the default starting date for all history runs is January 2000. However, the stock Easter[1] regressor with $d = 28$ has the value -0.152 in each March from the start of the series in 1992 through 2004 (with different values in March of 2005 and 2008). When this user-defined regressor is included in a model for a span of data that ends before March of 2005, seasonal differencing results in a regressor of zeros, causing a singularity in the regression matrix. As a result, when computing diagnostics (such as forecast error statistics) from the *history* spec of X-13A-S for a model with this regressor, the start date for the analysis was set to January of 2006.

graphed to form a graphical diagnostic for model selection. These plots will be called forecast error plots. The first model in a forecast plot is preferred if the direction of differences is mostly downward — in this case, the forecast errors are assumed to be smaller for the first model. Similarly, the second model is preferred if the direction of differences is mostly upward. However, when the differences do not appear to move in any specific direction or the differences seem to go in one direction for one forecast lead and in another direction for the other lead, a preferred model cannot be declared. The forecast error graph is discussed in Findley et. al. (1998). In our study we used log-transformed series to calculate forecast errors because the log transformation was needed for all models in the study.

4.2.3 Ratios of root mean square forecast errors

Records of forecast errors generated by the *history* spec of X-13A-S are also used to calculate root mean sums of squared forecast errors and ratios below.

For a time series Z_t, $1 \leq t \leq T$, and forecast lead $l \geq 1$ and forecast origin $1 \leq \tau \leq T$, we define $z_t = \log Z_t$ and let $z_{\tau+l|\tau}$ denote the forecast of $z_{\tau+l}$ calculated from z_t, $1 \leq t \leq \tau$, provided by the regARIMA model without calendar regressors when its parameter estimates are derived from z_t, $1 \leq t \leq \tau$. Let τ_0 denote the index of the initial forecast origin, then we define:

$$RMSE_l^{\varnothing} = \sqrt{(T - l - \tau_0 + 1)^{-1} \sum_{\tau=\tau_0}^{T-l} \left(z_{\tau+l} - z_{\tau+l|\tau}\right)^2}.$$

Forecasts of the original series, $Z_{\tau+l|\tau}$, are defined by exponentiating the forecasts $z_{\tau+l|\tau}$ of the logged series.

We define $RMSE_l$ as the corresponding root mean square forecast error for a lead in the model with calendar regressor(s). We further define $RMSE_{ratio} = RMSE_l^{\varnothing}/RMSE_l$ for $l = 1, 12$. If $RMSE_{ratio} > 1$ at both leads 1 and 12, then the model with calendar regressor(s) generally has smaller errors, thus providing support for the use of calendar regressor(s).

Let $RMSE_l^*$ be the root mean square forecast error for a lead in the model without stock Easter regressor (but this model can have any stock trading day regressor). Let $RMSE_l[w]$ denote the analogous value for the model with the stock Easter[w] regressor added. Then we define the ratio $RMSE_{ratio}^{E} = RMSE_l^*/RMSE_l[w]$ for $l = 1, 12$. The corresponding ratio $RMSE_{ratio}^{March} = RMSE_l^{March}/RMSE_l^{March}[w]$ for $l = 1, 12$, is defined using the March forecasts only, as described in Findley (2009). It should be noted that March is the month where the Easter effect regressors are always non-zero. If ratio values are greater than 1 at both leads 1 and 12, then the model with the stock Easter regressor is preferred because it provides root mean square forecast improvement at both leads.

4.3 Spectral diagnostics

Spectrum estimates of the last eight years of the (differenced and log-transformed) prior-adjusted original series, of the (differenced and log-transformed) seasonally adjusted series, of the irregulars modified for extreme values, and of the regARIMA model residuals were analyzed for the presence of visually significant (v.s.) trading day and seasonal peaks. In this paper, T1 and T2 refer respectively to the first and the second trading day frequencies (.348 and .432 cycles per month) in an estimated spectrum. It is expected that including a

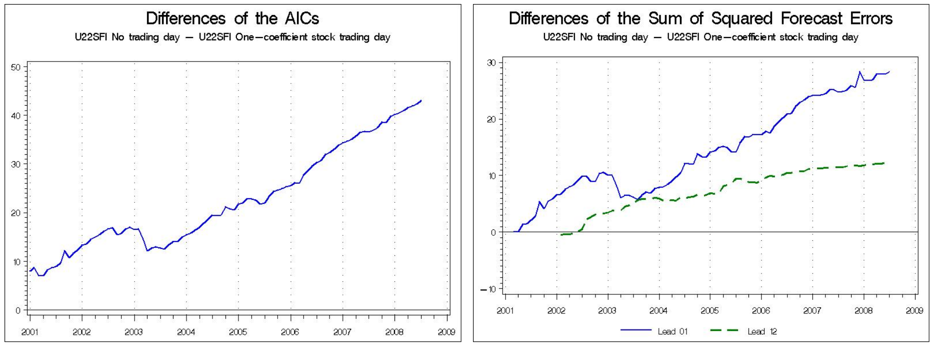

Figure 1: AIC history plot and forecast error plot (manufacturing inventory series 22SFI).

calendar regressor will eliminate v.s. peaks or will reduce the strength of such peaks (see Soukup and Findley (1999) and Section 6.1 of U.S. Census Bureau (2007) for more details).

4.4 Ljung-Box Q-statistics and ACF values

Numbers of significant Ljung-Box Q-statistics and of significant ACF values were also calculated. If a model with calendar regressor(s) has smaller numbers of significant Ljung-Box Q-statistics and ACF values than the corresponding model without calendar regressors, this usually supports the use of calendar regressor(s).

5. Example: model with the largest AICC difference

It is interesting to note that many models with one-coefficient stock trading day had very low AICC values compared with corresponding models without calendar regressors. The model which has the largest difference in AICC values from the model without calendar regressors is the model with the one-coefficient end-of-month stock trading day for the Inventories of Finished Goods for Paper Products (22SFI) series. The AICC difference for this series is $dAICC_N = AICC_N^A - AICC_N^B = 1637 - 1594 = 43$, and the AIC history plot in Figure 1 reveals that AIC differences lie in the positive range of approximately $(10, 40)$. The curve of AIC differences slopes upward over time, indicating that the preference for the model with the one-coefficient end-of-month stock trading day grows over time. There is a small jump downward after the year 2003, but in general the movement of the curve is upward.

The forecast error plot in Figure 1 shows that the direction of differences is mostly upward, which shows a preference for the second model, i.e., the model with the one-coefficient end-of-month stock trading day. Again, there is a small jump downward in lead 1 after the year 2003, but the general movement of lead 1 is upward, and its slope is even steeper than the slope of lead 12. In addition, $RMSE_{ratio} = 1.15$ for lead 1, and $RMSE_{ratio} = 1.07$ for lead 12. Because $RMSE_{ratio} > 1$ for both leads, then the model with the one-coefficient end-of-month stock trading day generally has smaller mean square errors at leads 1 and 12 (especially at lead 1).

Furthermore, after the inclusion of the one-coefficient end-of-month stock trading day regressor in the model, v.s. T1 and T2 peaks disappeared from the spectrum of the (differenced and log-transformed) seasonally adjusted series; v.s. T1 peak disappeared from the spectrums of the irregulars and of the regARIMA model residuals.

The inclusion of the one-coefficient end-of-month stock trading day regressor did not cause any change in the numbers of significant Ljung-Box Q-statistics and of ACF values. The model with the one-coefficient end-of-month stock trading day has no significant Ljung-Box Q-statistics and two significant ACF values but at different leads than in the model without calendar regressors.

In general, diagnostic results indicate that the model with the one-coefficient end-of-month stock trading day regressor fits best for the manufacturing 22SFI series.

6. Summary of diagnostic results

Tables 2 and 3 list all SSD and M3 series for which the minimum AICC criterion selected a model with calendar regressor(s). They also provide differences in AICC values, ratios of mean sums of squared forecast errors, and adjustment factors. A complete report with all diagnostic results is available upon request.

In general, diagnostic results tend to improve for those models with large AICC differences, but this is not always true. Furthermore, alternative models mentioned in footnotes under Tables 2 and 3 were checked for diagnostic results. They were considered to replace the corresponding main models only if their diagnostic results improved compared with those of the main models.

6.1 Retail and wholesale inventory series

Out of 7 retail and 20 wholesale series, the minimum AICC criterion selected a model with calendar regressor(s) for 5 retail and 10 wholesale series. According to Table 2, there are nine models with an end-of-month stock trading day regressor. The minimum AICC criterion also selected a model with the one-coefficient stock trading day with $d = 28$ for three wholesale series and a model with the stock Easter[1] with $d = 28$ for one wholesale series.

It is worthwhile to note particularities of the model with the one-coefficient end-of-month stock trading day for the p0b42330 series. This model seems not to work well for the early years approximately before 2004 (as revealed by diagnostic plots), but seems to work better in later years after 2004. For example, in the forecast error plot the lead 1 begins to slope downward in 2004. In the AIC history plot, the AIC difference becomes negative approximately after 2005. It looks like the trading day effect is present in the recent years only.

AICC results. Models for three retail series (p0b45200, p0b44100, p0b4423x) have AICC differences greater than 15. The wholesale series p0b42420 has the smallest AICC difference compared with other wholesale and retail series; therefore, it is not surprising that its AIC history plot lies mostly in the negative region. In fact, AIC history plots for two series (p0b42330, p0b42420) lie mostly in the negative region. However, for most series the AIC history plots slope upward and lie mostly in the positive region.

Forecasting. Forecast error plots are chaotic for some series (it is hard to say if a lead is generally moving upward or downward), but in some instances $RMSE_{ratio}$ reflects the general movement of a lead. Most series have $RMSE_{ratio} < 1$ for at least one lead (1 or 12), while four have $RMSE_{ratio} < 1$ for both and five have $RMSE_{ratio} > 1$ for both. In fact, only three series have agreement between the forecast error plots (both leads 1 and 12 are moving mostly upward) and $RMSE_{ratio}$ (p0b4423x, p0b42420, p0b42320).

Spectrum diagnostics. The addition of stock regressor(s) to the regARIMA model removed one v.s. peak or more from at least one spectrum for four series (p0b42300, p0b42330, p0b44100, p0b44800); conversely, adding stock regressor(s) to models caused

Table 2: Models selected for retail and wholesale inventory Service Sector series (Note: all alternative models mentioned in footnotes have AICC values that are significantly lower than AICC values of models without calendar regressors.) Differences in AICC values ($dAICC_N$, compared with the corresponding model without calendar regressors); ratios of root mean square forecast errors for selected models with calendar regressor(s); minimum and maximum of the seasonal (S), trading day (TD), and Easter effect (E) adjustment factors.

Series	$dAICC_N$	$RMSE_{ratio}$, $l = 1$	$RMSE_{ratio}$, $l = 12$	*min S*	*max S*	*min TD*	*max TD*	*min E*	*max E*
Retail inventory series									
Full end-of-month stock trading day									
p0b45200[2]	26.1	1.09	0.98	92.13	118.63	99.52	100.27	-	-
p0b45210[3]	3.6	0.99	0.99	91.12	121.83	99.65	100.27	-	-
One-coefficient end-of-month stock trading day									
p0b44100[4]	33.4	1.06	1.02	91.67	105.59	98.44	109.69	-	-
p0b4423x[4]	18.0	1.07	1.01	94.80	113.39	99.65	100.35	-	-
End-of-month stock Easter[8]									
p0b44800	6.9	0.99	0.99	90.20	113.13	-	-	99.23	100.42
Alternative models (with one-coefficient end-of-month stock trading day)									
p0b45200	23.5	1.09	1.00	92.07	118.65	99.70	100.30	-	-
p0b45210	2.2	1.01	1.00	91.10	121.88	99.86	100.14	-	-
Wholesale inventory series									
Full end-of-month stock trading day									
p0b42300[5]	1.9	0.99	1.01	98.45	101.33	99.80	100.12	-	-
p0b42330	2.0	0.97	0.99	94.41	105.44	99.57	100.64	-	-
p0b42420	1.5	1.02	1.02	94.01	109.65	99.25	100.64	-	-
One-coefficient end-of-month stock trading day									
p0b42350[6]	4.5	1.02	1.00	96.77	101.89	99.80	100.20	-	-
One-coefficient end-of-month stock trading day & end-of-month stock Easter[1]									
p0b42340	6.5	0.98	1.01	97.53	102.62	99.66	100.34	99.02	100.30
End-of-month stock Easter[8]									
p0b42320	4.9	1.02	1.02	97.08	103.69	-	-	98.95	100.58
One-coefficient stock trading day with $d = 28$									
p0b42360[7]	7.8	1.02	0.98	97.15	103.12	99.75	100.25	-	-
p0b42380[8]	4.6	1.02	0.99	97.77	102.30	99.83	100.17	-	-
p0b42480[9]	3.4	0.99	0.99	92.23	109.88	99.59	100.41	-	-
Stock Easter[1] *with $d = 28$*									
p0b42310	2.3	1.02	0.99	94.33	104.16	-	-	97.85	100.39

[2]The minimum AICC criterion chose this model over the model with the one-coefficient end-of-month stock trading day and over the model with the one-coefficient stock trading day with $d = 28$. The AICC value of the model with the full stock trading day with $d = 28$ is larger than the AICC value of this model.

[3]The minimum AICC criterion chose this model over the model with the one-coefficient end-of-month stock trading day and over the model with the one-coefficient stock trading day with $d = 28$.

[4]The minimum AICC criterion chose this model over the model with the full end-of-month stock trading day. The AICC values of the model with the full stock trading day with $d = 28$ and of the model with the one-coefficient stock trading day with $d = 28$ are larger than the AICC value of this model.

[5]The AICC value of the model with the full stock trading day with $d = 28$ is larger than the AICC value of this model. The AICC value of the model with the one-coefficient stock trading day with $d = 28$ is lower than the AICC value of this model by only 0.41.

[6]The AICC value of this model is by 1.92 lower than the AICC value of the model with the full stock trading day with $d = 28$.

[7]The minimum AICC criterion chose this model over the model with the full stock trading day with $d = 28$.

[8]The AICC value of this model is by 3.30 lower than the AICC value of the model with the full end-of-month stock trading day.

[9]The AICC value of the model with the full stock trading day with $d = 28$ is by 1.06 lower than the AICC value of this model.

the appearance of one v.s. peak or more in three series (p0b45210, p0b42420, p0b42320). In most cases, there was no change in the number of v.s. peaks detected.

Goodness of fit. For most series, adding stock calendar regressors either caused no change in the numbers of significant Ljung-Box Q-statistics and of ACF values or altered one/both of these numbers by one. For one series (p0b42300), adding stock calendar regressors increased the number of significant Ljung-Box Q-statistics by 4, and for another series (p0b42330) increased both numbers of significant Ljung-Box Q-statistics and of significant ACF values by 2. The final model for the wholesale p0b42360 series had two fewer significant Ljung-Box Q-statistics than the model without calendar regressors, and the final model for the wholesale p0b42480 series had four fewer significant Ljung-Box Q-statistics.

Adjustment factors. Trading day factors have the largest range, from 98.44 to 109.69, for the retail p0b44100 series, for which the model with the one-coefficient end-of-month stock trading day was chosen (seasonal factors range from 91.67 to 105.59). Trading day factors have the smallest range, from 99.80 to 100.12, for the wholesale p0b42300 series, for which the model with the full end-of-month stock trading day was selected (seasonal factors range from 98.45 to 101.33). However, the retail p0b45210 series with the alternative model (discussed below) has trading day factors with a smaller range, from 99.86 to 100.14 (seasonal factors range from 91.10 to 121.88). Holiday factors have the largest range, from 97.85 to 100.39, for the wholesale p0b42310 series, whose final model is the model with the stock Easter[1] with $d = 28$ (seasonal factors range from 94.33 to 104.16). Holiday factors have the smallest range, from 99.23 to 100.42, for the retail p0b44800 series, for which the model with the end-of-month stock Easter[8] was chosen (seasonal factors range from 90.20 to 113.13).

6.1.1 Alternative models for retail inventory series

For the retail p0b45200 series, the model with the one-coefficient end-of-month stock trading day does not eliminate the v.s. T2 peak from the spectrum of the residuals, and the strength of this peak actually increases but is still lower than the peak from the model with the full end-of-month stock trading day. Also, this alternative model improves forecast diagnostics — now both leads 1 and 12 in the forecast error plot are sloping mostly downward, and $RMSE_{ratio} > 1$ for both leads 1 and 12. Hence, it was chosen as the final model.

For the retail p0b45210 series, the inclusion of the full end-of-month stock trading day resulted in the appearance of a v.s. S2 peak in the spectrum of residuals. The alternative model with the one-coefficient end-of-month stock trading day does not have this problem; also, its forecast error plot looks better, and $RMSE_{ratio} > 1$ for both leads. It was chosen as the final model as well.

6.2 M3 series

Out of 96 manufacturing inventory series, the minimum AICC criterion selected a model with calendar regressor(s) for 40 series. According to Table 3, the most common choice among the 40 series was the one-coefficient end-of-month stock trading day regressor (12 in total). There are also many models with a stock trading day with $d = 28$ (11 in total). A model with a stock Easter regressor was selected for 17 series.

AICC results. Models for six series (31ATI, 22SFI, 26SFI, 27SFI, 31SFI, 26SMI) have AICC differences greater than 15, and five of them have the one-coefficient end-of-month stock trading day regressor. The model for the 34CTI series has the smallest AICC difference compared with other manufacturing series. As a result, its diagnostic results are rather unsatisfactory. For most models the AIC history plot slopes upward and lies mostly in the positive region, but for three series (34CTI, 25CTI, 32SWI) these plots lie mostly in the negative region.

Forecasting. Forecast error plots are satisfactory only for eight series (22ATI, 22SFI, 27SFI, 25SMI, 26SMI, 32SFI, 36SFI, 12BTI), which also have $RMSE_{ratio} > 1$ for both leads. Most series have $RMSE_{ratio} < 1$ for only one lead (lead 1 or 12). Fifteen series have $RMSE_{ratio} > 1$ for both leads, and four have $RMSE_{ratio} < 1$ for both leads (34CTI, 33SFI, 34SMI, 33ATI).

Spectrum diagnostics. The addition of stock regressor(s) to the regARIMA model removed one v.s. peak or more from a spectrum for 9 series, while for fourteen series it caused the appearance of one v.s. peak or more in a spectrum. For two series, adding stock regressor(s) caused the removal of one or two v.s. peaks and the appearance of one v.s. peak (26SMI, 39SMI). For fifteen series there was no change in the number of v.s. peaks.

Goodness of fit. Either there was no change in the number of significant Ljung-Box Q-statistics and ACF values or the number of significant Ljung-Box Q-statistics and/or of significant ACF values was altered by 1. For one series (31ATI), adding stock calendar regressor(s) increased the number of significant Ljung-Box Q-statistics by 8, and for another series (34KTI) it reduced the number of significant Ljung-Box Q-statistics by 3. Four series (11BTI, 11SFI, 12ATI, 32SFI) had the number of significant ACF values increase by 2, while for other four series (27SFI, 25CTI, 27SWI, 16SWI) the number of significant ACF values was reduced by 2.

Adjustment factors. Trading day factors have the largest range, from 98.27 to 102.01, for the 36ATI series, whose final model is the model with the full end-of-month stock trading day and the end-of-month stock Easter[15] (seasonal factors range from 88.75 to 107.41). Trading day factors have the smallest range, from 99.86 to 100.14, for the 25SMI series, for which the model with the one-coefficient end-of-month stock trading and the end-of-month stock Easter[8] was selected (seasonal factors range from 98.40 to 101.14). Holiday factors have the largest range, from 97.73 to 103.56, for the 24SWI series, for which the model with the stock Easter[15] with $d = 28$ was chosen (seasonal factors range from 90.17 to 105.07). Holiday factors have the smallest range, from 99.77 to 100.41, for the 25CTI series, whose final model is the model with the end-of-month stock Easter[8] (seasonal factors range from 94.63 to 103.90).

6.2.1 Alternative model for one M3 series

For the 12ATI series, the main model with the full stock trading day with $d = 28$ has $RMSE_{ratio} < 1$ for lead 12, but the alternative model with the full end-of-month stock trading day has $RMSE_{ratio} > 1$ for both leads and a relatively better AIC history plot. Therefore, the alternative model was chosen as the final model.

6.3 Side notes on models with different stock Easter regressors

For one retail, one wholesale, and five manufacturing inventory series the minimum AICC criterion selected more than one model with the same number of parameters but different stock Easter regressors. For the selection of a final model from among these non-nested models, we compared AICC values and diagnostic results. For example, for

Table 3: Models selected for M3 series (Note: all alternative models mentioned in footnotes have AICC values that are significantly lower than AICC values of models without calendar regressors.) Differences in AICC values ($dAICC_N$, compared with the corresponding model without calendar regressors); ratios of root mean square forecast errors for selected models with calendar regressor(s); minimum and maximum of the seasonal (S), trading day (TD), and Easter effect (E) adjustment factors.

Series	$dAICC_N$	$RMSE_{ratio}$, $l=1$	$RMSE_{ratio}$, $l=12$	*min S*	*max S*	*min TD*	*max TD*	*min E*	*max E*
				Full end-of-month stock trading day					
22ATI[10]	12.3	1.05	1.02	97.49	101.99	99.74	100.18	-	-
34CTI	0.7	0.99	0.99	89.66	106.21	99.26	100.47	-	-
				Full end-of-month stock trading day & end-of-month stock Easter[15]					
31ATI[10]	29.7	1.03	0.99	98.20	101.99	99.62	100.27	99.68	100.36
36ATI[11]	13.8	1.02	0.99	88.75	107.41	98.27	102.01	97.76	102.59
				One-coefficient end-of-month stock trading day					
11BTI[12]	9.9	1.03	1.00	91.55	105.00	99.55	100.46	-	-
15SWI	2.5	0.99	1.00	92.62	106.36	99.75	100.25	-	-
16SFI[13]	5.5	1.01	1.00	92.53	107.49	99.57	100.44	-	-
22SFI[12]	43.0	1.15	1.07	96.81	103.06	99.36	100.64	-	-
24SFI	1.9	1.01	0.99	92.36	104.91	99.52	100.48	-	-
26SFI[14]	15.7	1.01	1.02	95.67	104.31	99.73	100.27	-	-
27SFI[12]	30.1	1.13	1.06	94.73	104.46	99.40	100.61	-	-
31SFI[12]	42.5	1.28	1.01	96.64	103.38	99.32	100.69	-	-
31SMI[15]	10.6	0.99	1.00	96.59	104.23	99.67	100.33	-	-
31SWI[15]	12.1	0.99	1.01	97.03	101.75	99.73	100.27	-	-
33SFI	2.7	0.98	0.99	97.67	104.82	99.83	100.17	-	-
36CTI[16]	11.7	1.01	0.99	94.90	104.28	99.28	100.73	-	-
				One-coefficient end-of-month stock trading day & end-of-month stock Easter[8]					
11SFI	8.9	0.97	1.00	93.50	106.43	99.76	100.24	99.31	100.38
24ATI	7.1	1.02	0.99	91.83	105.16	99.68	100.32	99.28	101.34
25SMI	3.7	1.01	1.00	98.40	101.14	99.86	100.14	99.35	100.36
26SMI[16]	15.6	1.03	1.00	98.92	101.09	99.71	100.29	99.53	100.87
				End-of-month stock Easter[1]					
34SMI	2.0	0.99	0.99	97.88	103.31	-	-	99.68	101.06
				End-of-month stock Easter[8]					
25CTI	5.1	1.02	0.99	94.63	103.90	-	-	99.77	100.41
33ATI	2.7	0.98	0.99	90.83	110.91	-	-	99.20	100.44

[10]The minimum AICC criterion chose the model with the full end-of-month stock trading day over the model with the one-coefficient end-of-month stock trading day and over the model with the one-coefficient stock trading day with $d = 28$. The AICC value of the model with the full stock trading day with $d = 28$ is larger than the AICC value of the model with the full end-of-month stock trading day.

[11]The AICC value of the model with the full stock trading day with $d = 28$ is larger than the AICC value of the model with the full end-of-month stock trading day.

[12]The minimum AICC procedure chose this model over the model with the full end-of-month stock trading day. The AICC values of the model with the full stock trading day with $d = 28$ and of the model with the one-coefficient stock trading day with $d = 28$ are larger than the AICC value of this model.

[13]The AICC value of the model with the one-coefficient stock trading day with $d = 28$ is larger than the AICC value of this model.

[14]The minimum AICC procedure chose this model over the model with the full end-of-month stock trading day. The AICC value of the model with the full stock trading day with $d = 28$ is larger than the AICC value of this model.

[15]The minimum AICC procedure chose this model over the model with the full end-of-month stock trading day. The AICC value of the model with the one-coefficient stock trading day with $d = 28$ is larger than the AICC value of this model.

[16]The minimum AICC procedure chose the model with the one-coefficient end-of-month stock trading day over the model with the full end-of-month stock trading day.

Table 3: Models selected for M3 series (Note: all alternative models mentioned in footnotes have AICC values that are significantly lower than AICC values of models without calendar regressors.) Differences in AICC values ($dAICC_N$, compared with the corresponding model without calendar regressors); ratios of root mean square forecast errors for selected models with calendar regressor(s); minimum and maximum of the seasonal (S), trading day (TD), and Easter effect (E) adjustment factors.

35SFI	9.5	1.02	0.99	91.58	106.84	-	-	99.04	100.53
39SMI	2.8	0.99	1.00	98.32	101.68	-	-	99.64	100.67
End-of-month stock Easter[15]									
21SFI	4.5	1.01	1.00	93.31	108.19	-	-	98.84	101.33
Full stock trading day with $d = 28$									
12ATI[17]	9.4	1.03	0.99	96.57	104.64	99.47	100.50	-	-
27SWI	6.1	0.99	1.04	93.06	105.56	99.58	100.51	-	-
31CTI	3.9	0.99	1.00	95.96	103.76	99.64	100.38	-	-
32SWI	1.3	1.02	0.99	96.28	102.22	99.64	100.33	-	-
34KTI[18]	11.5	0.99	1.00	97.41	102.24	99.74	100.47	-	-
35ATI	2.5	1.00	0.99	93.20	103.74	99.65	100.34	-	-
Full stock trading day with $d = 28$ *& stock Easter*[8] *with* $d = 28$									
32SFI	10.8	1.06	1.02	94.59	104.84	99.80	100.25	99.00	100.39
One-coefficient stock trading day with $d = 28$									
11ATI[19]	7.2	1.03	1.00	92.14	109.23	99.43	100.57	-	-
16SWI	6.8	1.01	0.99	95.04	104.86	99.19	100.81	-	-
36SFI[19]	2.1	1.01	1.00	90.11	105.34	99.64	100.37	-	-
One-coefficient stock trading day with $d = 28$ *& stock Easter*[1] *with* $d = 28$									
12BTI[20]	11.2	1.24	1.04	90.61	108.78	99.70	100.30	97.66	100.42
Stock Easter[1] *with* $d = 28$									
15SFI	2.2	1.01	0.97	90.48	110.01	-	-	98.41	100.29
34BTI	3.3	1.04	0.99	85.62	110.00	-	-	99.33	103.82
Stock Easter[15] *with* $d = 28$									
24SWI	5.1	1.02	1.01	90.17	105.07	-	-	97.73	103.56
Alternative model (with full end-of-month stock trading day)									
12ATI	4.6	1.01	1.01	96.64	104.54	99.59	100.39	-	-

the retail inventory series p0b42320 the AICC values of the model with the end-of-month stock Easter[1] and of the model with the stock Easter[8] with $d = 28$ are lower than the AICC value of the model with the end-of-month stock Easter[8], but we chose the latter model because the first two models have relatively worse diagnostic results. Because the end-of-month Easter[8] tends to be more common in flow series (in accordance with past experience), we chose it for the manufacturing 11SFI and 24ATI series instead of the end-of-month stock Easter$[w]$ with $w = 1, 15$.

6.4 Easter coefficient and RMSE ratio values for models with a stock Easter

According to Table 4, models with a stock Easter regressor for six series have all four ratios of RMSE values greater than 1. In models with a stock Easter regressor, for nine

[17] The AICC value of the model with the full end-of-month stock trading day is larger than the AICC value of this model by more than 2.

[18] The AICC values of the model with the full end-of-month stock trading day, the model with the one-coefficient end-of-month stock trading day, and the model with the one-coefficient stock trading day with $d = 28$ are larger than the AICC value of this model by more than 2.

[19] The minimum AICC procedure chose this model over the model with the full stock trading day with $d = 28$.

[20] The AICC value of the model with the full end-of-month stock trading day is larger than the AICC value of the model with the one-coefficient stock trading day with $d = 28$ by more than 2.

Table 4: Easter coefficient and $RMSE$ ratio values for models with a stock Easter regressor.

Series	Stock day	w	Stock TD regressors?	$RMSE^{E}_{ratio}$, $l=1$	$RMSE^{March}_{ratio}$, $l=1$	$RMSE^{E}_{ratio}$, $l=12$	$RMSE^{March}_{ratio}$, $l=12$	Easter reg. coef.
Retail inventory series								
p0b44800	end-of-month	8	-	0.99	0.99	0.99	1.03	-0.01
Wholesale inventory series								
p0b42340	end-of-month	1	One-coef.	0.96	0.71	1.01	1.12	-0.01
p0b42320	end-of-month	8	-	1.02	1.27	1.02	1.12	-0.02
p0b42310	$d = 28$	1	-	1.02	2.23	0.99	1.03	-0.03
Manufacturing inventory series								
31ATI	end-of-month	15	Full	1.02	1.06	0.99	0.99	0.01
36ATI	end-of-month	15	Full	1.01	1.16	0.99	0.97	0.05
11SFI	end-of-month	8	One-coef.	1.02	1.03	0.99	0.99	-0.01
24ATI	end-of-month	8	One-coef.	1.01	1.24	1.00	0.97	0.02
25SMI	end-of-month	8	One-coef.	1.01	1.11	1.00	1.00	-0.01
26SMI	end-of-month	8	One-coef.	1.01	1.20	0.99	0.98	0.01
34SMI	end-of-month	1	-	0.99	0.96	0.99	0.94	0.01
25CTI	end-of-month	8	-	1.02	1.13	0.99	0.99	0.01
33ATI	end-of-month	8	-	0.98	0.53	0.99	0.90	-0.01
35SFI	end-of-month	8	-	1.02	1.48	0.99	1.05	-0.01
39SMI	end-of-month	8	-	0.99	1.11	1.00	1.01	0.01
21SFI	end-of-month	15	-	1.01	1.09	1.00	1.01	0.02
32SFI	$d = 28$	8	Full	1.05	1.11	1.01	1.02	-0.01
12BTI	$d = 28$	1	One-coef.	1.24	2.54	1.05	9.45	-0.03
15SFI	$d = 28$	1	-	1.01	1.70	0.97	0.88	-0.02
34BTI	$d = 28$	1	-	1.04	2.01	0.99	1.13	0.04
24SWI	$d = 28$	15	-	1.02	1.14	1.01	1.04	0.06

series the root mean square error improvement is greater for the March only forecasts at both leads 1 and 12. In models with a stock Easter regressor, for ten series the root mean square error improvement is greater for the March only forecasts at one lead. The models for the manufacturing 34SMI and 33ATI series have all four ratios of RMSE values less than 1. The greatest ratio is $RMSE^{March}_{12}/RMSE^{March}_{12}[w] = 9.45$ of the model with the one-coefficient stock trading day with $d = 28$ and the stock Easter[1] with $d = 28$ for the manufacturing 12BTI series, but this ratio is based on relatively few data points and is rather an outlier among other ratios, which are quite close to 1.

It is interesting to note that all stock Easter regressor coefficients are negative for the retail and wholesale inventory series, whereas the majority (10 of 17) of manufacturing inventory series have positive coefficients.

7. Conclusions

Most of the retail and wholesale inventory and M3 series with significant stock calendar regressors have final models with the one-coefficient end-of-month stock trading day regressor. This agrees with the results of Findley and Monsell (2009), who discovered that the use of a one-coefficient regressor is likely to significantly increase the number of manufacturing series in which such effects are identified.

Also, with the final regressor choices, stock Easter effects were identified in four Service Sector inventory series and seventeen M3 inventory series. The Easter effect interval lengths of the selected regressors included all of three of the lengths considered ($w = 1, 8, 15$). Among these 21 series, there were six for which a stock regressor for the 28th day of the month was preferred. In each case, this might indicate that enough respondents could only supply inventory numbers for some day prior to the end of the month to cause the selected regressor to provide a more representative Easter effect than the end-of-month regressor.

Acknowledgements. The authors are grateful to David F. Findley and William R. Bell of U.S. Census Bureau for helpful comments and suggestions that substantially improved this work.

REFERENCES

Bell, W. R. (1984), "Seasonal Decomposition of Deterministic Effects," *Statistical Research Division Research Report No. 84/01*, U.S. Census Bureau. Available online at http://www.census.gov/srd/papers/pdf/rr84-1.pdf.

Bell, W. R. (1995), "Correction to Seasonal Decomposition of Deterministic Effects," *Statistical Research Division Research Report No. 95/01*, U.S. Census Bureau. Available online at http://www.census.gov/srd/papers/pdf/rr95-01.pdf.

Burnham, K. P., and Anderson, D. R. (2004), "Multimodal Inference: Understanding AIC and BIC in Model Selection," *Sociological Methods & Research*, 33, 261–304.

Findley, D. F. (2009), "Stock Series Holiday Regressors Generated By Flow Series Holiday Regressors," *Statistical Research Division Research Report Series (Statistics #2009-04)*, U.S. Census Bureau. Available online at http://www.census.gov/srd/papers/pdf/rrs2009-04.pdf.

Findley, D. F., and Monsell, B. C. (2009), "Modeling Stock Trading Day Effects Under Flow Day-of-Week Effect Constraints," *Journal of Official Statistics* (to appear). Available online at http://www.census.gov/ts/papers/findleymonselljos2008.pdf.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B. C. (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics*, 16, 127–177 (with discussion).

Hurvich, C., and Tsai, C. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297–307.

Monsell, B. C. (2007), "Issues in Modeling and Adjusting Calendar Effects in Economic Time Series," Proceedings of the Third International Conference on Establishment Surveys [CD-ROM], Alexandria, VA: American Statistical Association. Available online at http://www.census.gov/ts/papers/ices2007bcm.pdf.

M3 (2009), "2009 Instruction Manual For Reporting in Monthly Survey M3 Manufacturers' Shipments, Inventories, and Orders. U.S. Census Bureau." Available online at http://www.census.gov/indicator/www/m3/contactinfo/instructionman.pdf.

Soukup, R. J., and Findley, D. F. (1999), "On the Spectrum Diagnostics Used by X-12-ARIMA to Indicate the Presence of Trading Day Effects after Modeling or Adjustment," *1999 Proceedings of the Business and Economic Statistics Section*, pp. 144–149, Alexandria, VA: American Statistical Association. Available online at http://www.census.gov/ts/papers/rr9903s.pdf.

Taniguchi, M., and Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference of Time Series*, New York: Springer-Verlag.

U.S. Census Bureau (2009), "Economic Indicators." Available online at http://www.census.gov/cgi-bin/briefroom/BriefRm.

U.S. Census Bureau (2007), "X-12-ARIMA Reference Manual Version 0.3." Available online at http://www.census.gov/ts/x12a/v03/x12adocV03.pdf.